Industrial Internship Report on

Quality Prediction in a Mining Process

Prepared by

Kartavya Rakeshkumar Patel

Executive Summary
This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT). This internship was focused on a project/problem statement provided by UCT. We had to finish the project including the report in 6 weeks' time. My project was Quality Prediction in a Mining Process. This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship.

TABLE OF CONTENTS

1 Preface

Over the past six weeks at Uniconverge Technologies, I embarked on an exhilarating journey into the heart of the data-driven world. From understanding the transformative power of Big Data on businesses to unraveling the distinctions between Data Scientists and Data Analysts, I discovered the foundation of modern data insights.

Venturing into Probability and Statistics, I honed the essential skills of a data enthusiast. This foundation paved the way for a deep dive into Machine Learning, where I explored linear functions, optimization techniques, and the fascinating world of different learning types.

Week-5 saw me ascend the success ladder to the corporate world, gaining a clear vision of how to navigate this dynamic realm. I uncovered the potential earnings of a Data Engineer and equipped myself with Data Science interview questions, ready to seize future opportunities.

About need of relevant Internship in career development:

Relevant internships are the launchpads of career development. They provide hands-on experience, insights, and networking opportunities that propel me forward in my chosen field. Without them, it's like trying to climb a mountain without proper gear.

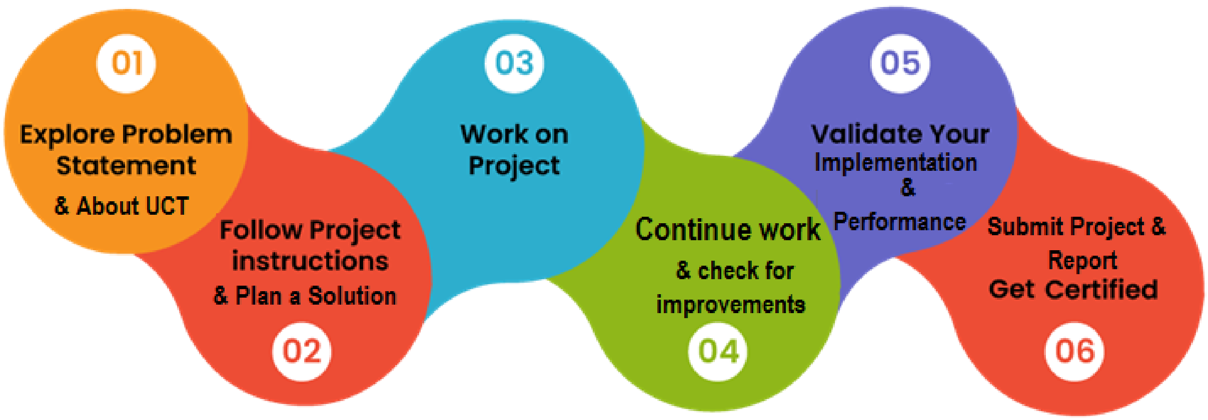
Brief about Your project/problem statement.

This project aims to predict silica impurity levels in the iron ore concentrate of a mining flotation plant. By providing early information, engineers can optimize the process, reduce waste, and take proactive actions. The project's outcomes will enhance efficiency and sustainability in the mining industry. Exploring real industrial data and help manufacturing plants to be more efficient.

Opportunity given by USC/UCT:

USC/UCT provided me with immersive internships, allowing hands-on experience, skill development, industry connections, and a platform to kickstart a promising career journey.

How Program was planned:



Your Learnings and overall experience:

My internship at USC/UCT has been a profound learning experience. From gaining practical skills and insights into the industry to building a valuable professional network, the opportunity they provided was invaluable. The hands-on experience has deepened my understanding of the field, and the challenges I faced have honed my problem-solving abilities. I'm grateful for the chance to apply classroom knowledge in real-world scenarios, which has boosted my confidence and enthusiasm for my chosen career path. Overall, this internship has been a pivotal step in my development, and I'm excited to take the skills and experiences I've gained here into my future endeavors.

Thanks to all, who have helped me directly or indirectly.

To my juniors and peers, embrace every opportunity, learn from challenges, and nurture your curiosity. Together, we'll shape the future with our passion, determination, and continuous growth. Let's support each other, innovate, and leave our mark on the world. Success awaits, so let's chase it!

2 Introduction

2.1 About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

For developing its products and solutions it is leveraging various **Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end** etc.

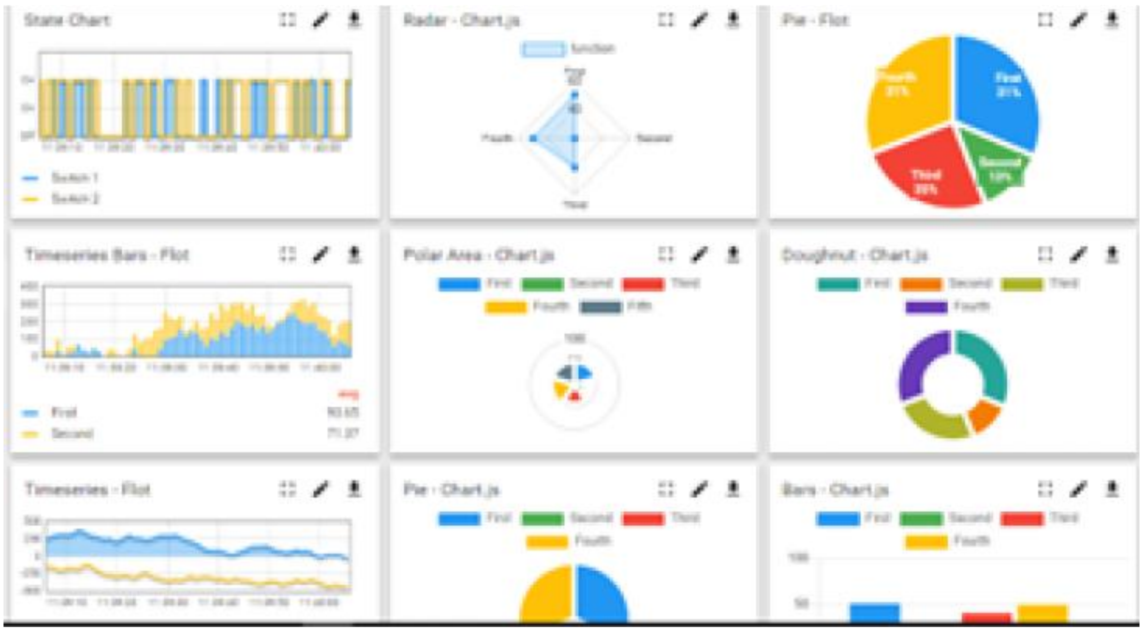
i. UCT IoT Platform ( )

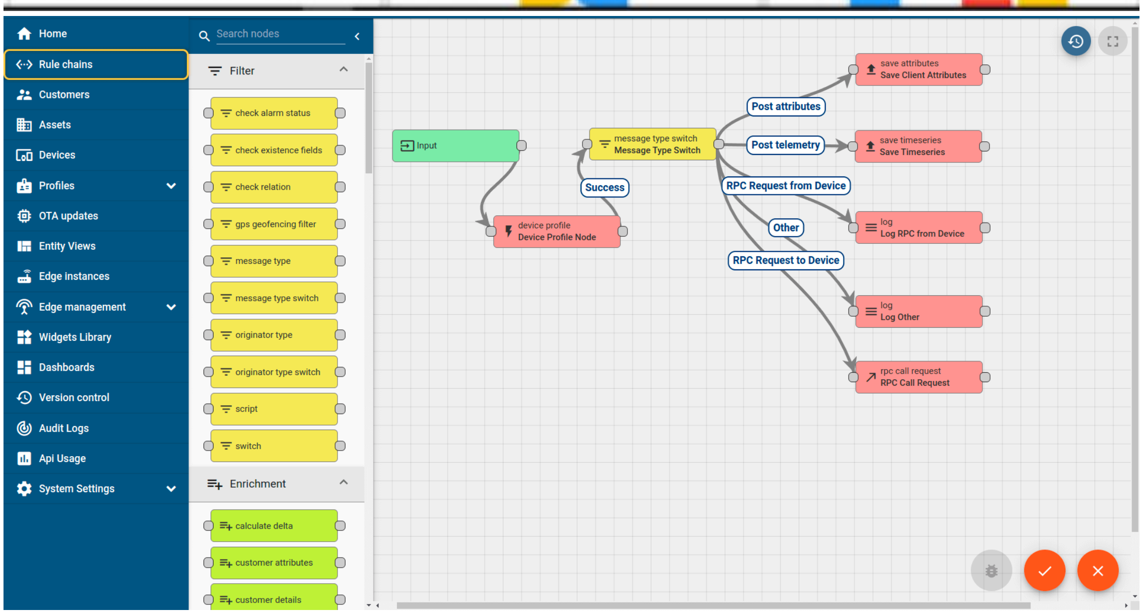
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
- It supports both cloud and on-premises deployments.

It has features to

- Build Your own dashboard
- Analytics and Reporting
- Alert and Notification
- Integration with third party application(Power BI, SAP, ERP)
- Rule Engine





ii. Smart Factory Platform (FACTORY WATCH)

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

- with a scalable solution for their Production and asset monitoring
- OEE and predictive maintenance solution scaling up to digital twin for your assets.
- to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
- A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

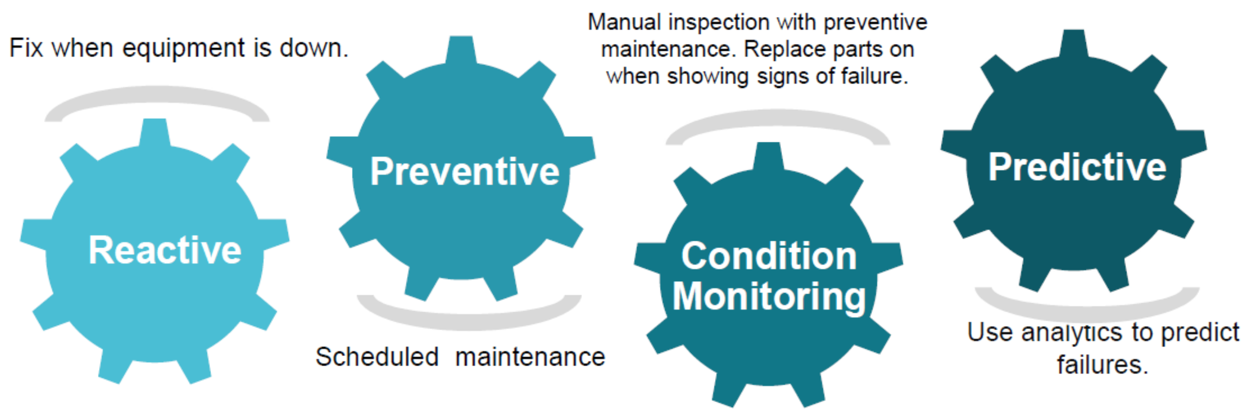
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i

iii. LoRaWAN based Solution

UCT is one of the early adopters of LoRAWAN teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

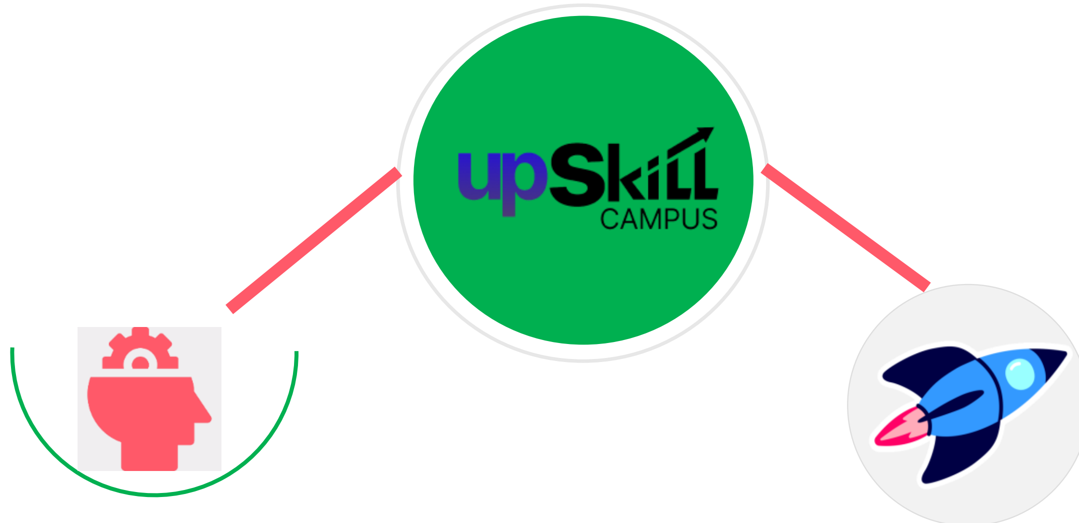
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

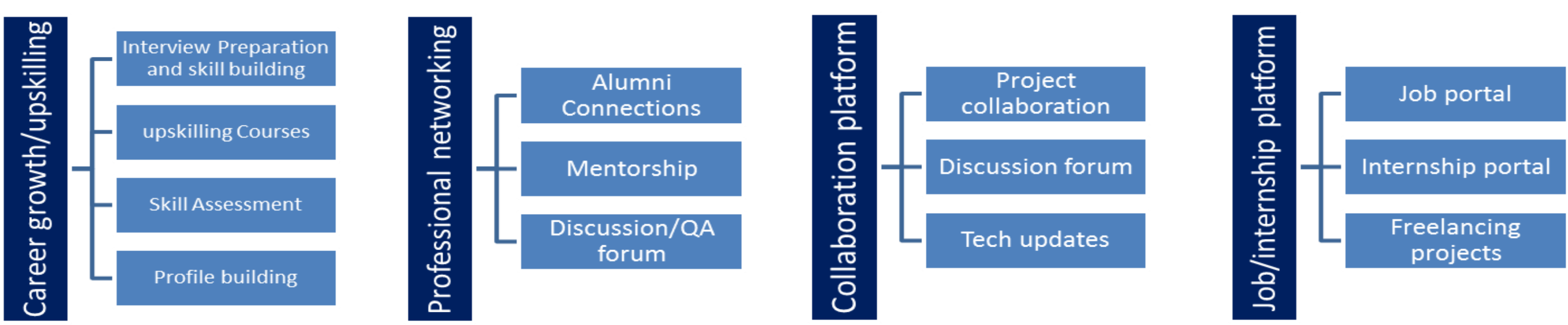
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

upSkill Campus aiming to upskill 1 million learners in next 5 year

https://www.upskillcampus.com/



2.3 The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

2.4 Objectives of this Internship program

The objective for this internship program was to

☛ get practical experience of working in the industry.

☛ to solve real world problems.

☛ to have improved job prospects.

☛ to have Improved understanding of our field and its applications.

☛ to have Personal growth like better communication and problem solving.

2.5 Reference

[1] Edunet Foundation

2.6 Glossary

Terms	Acronym
Big Data	B**D**
Machine Learning	M**L**
Data Engineer	DE
Data Scientist	DS
Artificial Intelligence	AI

3 Problem Statement

It is not always easy to find databases from real world manufacturing plants, specially mining plants. This database comes from one of the most important parts of a mining process: a flotation plant.

The main goal is to use this data to predict how much impurity is in the ore concentrate. As this impurity is measured every hour, if we can predict how much silica (impurity) is in the ore concentrate, we can help the engineers, giving them early information to take actions (empowering!).

Hence, they will be able to take corrective actions in advance (reduce impurity, if it is the case) and also help the environment (reducing the amount of ore that goes to tailings as you reduce silica in the ore concentrate).

The mining industry faces the challenge of predicting the quality of output materials accurately and in real-time. The existing manual methods are time-consuming, prone to errors, and lack the ability to account for dynamic process variations.

This project aims to develop an advanced data-driven solution that leverages machine learning to predict material quality, enhancing operational efficiency and ensuring the production of high-quality end products.

4 Existing and Proposed solution

In the domain of quality prediction in mining processes, existing methods often rely on manual sampling and laboratory analysis. However, these approaches have limitations. They are time-consuming, providing delayed insights. Additionally, they might not capture real-time variations, leading to suboptimal decision-making.

Our proposed solution involves the deployment of advanced data analytics and machine learning techniques to predict the quality of the mining process in real-time. We aim to develop predictive models that can anticipate variations in quality and identify potential issues before they impact the final product.

Transitioning from manual sampling to predictive analytics promises to revolutionize the mining process, fostering efficiency, cost savings, and enhanced product quality.

4.1 Code submission (Github link)

- https://github.com/kart12368/upskillcampus/blob/main/QualityPredictionInaMiningProcess.ipynb

4.2 Report submission (Github link)

- https://github.com/kart12368/upskillcampus/blob/main/QualityPredictionInaMiningProcess_Kartavya_USC_UCT.pdf

5 Proposed Design/ Model

- **Feature Engineering:** Feature engineering plays a pivotal role in enhancing model performance. Domain knowledge guides the selection of features that hold predictive power. Extracting meaningful features, combining attributes, and normalizing data create a feature set that fuels the machine learning model's accuracy.

- **Model Selection**: A range of machine learning algorithms, including decision trees, random forests, and gradient boosting, are evaluated based on the project's objectives and dataset characteristics. Cross-validation techniques are employed to determine the algorithm that yields the best predictive performance.

- **Model Training and Tuning**: The selected algorithm is trained on a portion of the dataset, and hyperparameter tuning is conducted to optimize model parameters. This phase involves adjusting parameters to minimize errors and enhance the model's ability to generalize to unseen data.

- **Model Evaluation**: The trained model is evaluated using metrics such as accuracy, precision, recall, and F1-score. The evaluation stage ensures that the model's performance aligns with project goals and meets the desired level of accuracy in predicting the quality of mining output.

5.1 High Level Diagram (if applicable)

Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.2 Low Level Diagram (if applicable)

5.3 Interfaces (if applicable)

Update with Block Diagrams, Data flow, protocols, FLOW Charts, State Machines, Memory Buffer Management.

6 Performance Test

In evaluating the real-world applicability of the proposed Quality Prediction in a Mining Process solution, rigorous performance testing was conducted to ensure its viability within industrial constraints. Several key constraints were identified, and strategies were implemented to address them, resulting in insightful test outcomes:

- **Memory and Processing Speed:** One critical constraint is the available memory and processing speed within the mining environment. To address this, model optimization techniques were applied during the design phase. Feature selection, dimensionality reduction, and algorithm optimization were employed to minimize memory usage and enhance processing speed.

- **Accuracy and Durability:** The accuracy of the predictive model and its ability to maintain performance over time were paramount. To handle this, a combination of robust feature engineering and continuous monitoring was implemented. This approach ensures that the model remains accurate and adaptable, accounting for variations in the mining process.

- **Recommendations for Untested Constraints:** Constraints such as power consumption and durability were addressed indirectly by ensuring model efficiency and monitoring. In cases where constraints were not directly tested, it's important to acknowledge their potential impact. Recommendations include continuous monitoring of power consumption and regular model maintenance to ensure durability.

6.1 Test Plan/ Test Cases

6.2 Test Procedure

Data sets were prepared to reflect different operational scenarios. Tests were conducted on a collab GPU environment. Data inputs were adjusted to assess the solution's performance under various conditions.

6.3 Performance Outcome

• Is it possible to predict % Silica Concentrate every minute?

Ans. Based on the analysis of the unique time differences between consecutive records in the dataset, it appears that there isn't a consistent 1-minute interval for the % Silica Concentrate records. The irregular time intervals, including 1 hour and 1 day 7 hours intervals, suggest that the dataset does not support predicting % Silica Concentrate every minute with this specific data.

Therefore, based on the information provided, it does not seem possible to predict % Silica Concentrate every minute using this dataset. The time intervals between the records are not suitable for minute-by-minute predictions of % Silica Concentrate.

• How many steps (hours) ahead can we predict % Silica in Concentrate? This would help engineers to act in predictive and optimized way, mitigating the % of iron that could have gone to tailings.

Ans. The dataset does not provide a consistent and reliable time interval for predicting % Silica in Concentrate. The maximum time difference (gap) between consecutive records in the dataset is approximately 319 hours, which corresponds to about 13 days. This means that, based solely on the available data, it is challenging to predict % Silica in Concentrate with a high degree of reliability beyond a lead time of approximately 13 days. Engineers should consider this limitation when making predictive and optimized decisions, particularly if they aim to mitigate the % of iron that could have gone to tailings over longer timeframes.

• Is it possible to predict % Silica in Concentrate without using % Iron Concentrate column (as they are highly correlated)?

Ans. Based on the observed significant increase in the Mean Squared Error (MSE) when the '% Iron Concentrate' column (which is highly correlated with '% Silica Concentrate') was removed from the model, it appears that predicting '% Silica Concentrate' without using the '% Iron Concentrate' column as a feature is challenging and may not be effective. The increase in MSE indicates that the model's predictive performance is adversely affected when this correlated feature is excluded.

In practice, when two features are highly correlated, they often contain similar information about the target variable. Removing one of the highly correlated features can result in a loss of essential information for prediction, leading to decreased model performance. Therefore, in this specific case, it seems that the '% Iron Concentrate' column provides valuable information for predicting '% Silica Concentrate,' and excluding it from the model may not yield accurate predictions.

To achieve the best predictive performance for '% Silica Concentrate,' it is advisable to include the '% Iron Concentrate' column as a feature in the model due to its correlation with the target variable.

7 My learnings

Throughout the course of this project on Quality Prediction in a Mining Process, I have accumulated a wealth of insights and experiences that have significantly contributed to my personal and professional growth.

- **Technical Mastery:** Navigating the intricacies of data preprocessing, feature engineering, and predictive modeling has elevated my technical prowess. I've gained a comprehensive understanding of machine learning algorithms, empowering me to transform raw data into actionable insights.
- **Problem-Solving and Adaptability:** Working on this project exposed me to the complexities and uncertainties inherent in real-world data scenarios. Navigating through challenges, troubleshooting issues, and adapting my approach have sharpened my problem-solving abilities and equipped me to handle diverse situations.
- **Industry Relevance:** Understanding the practical implications of quality prediction in a mining process has provided me with a glimpse into the industrial applications of data science. This exposure has broadened my perspective on how data-driven insights contribute to operational efficiency and informed decision-making.

This project has been an enriching journey that encapsulates both technical expertise and personal development. The practical skills acquired, coupled with the invaluable insights gained, will undoubtedly shape my career trajectory. I'm excited to carry these learnings forward, applying them to future endeavors and making a positive impact in the dynamic landscape of data science and industry.

8 Future work scope

- **Integration of External Data Sources**: Incorporating data from external sources, such as environmental factors or market demand, could provide a more comprehensive understanding of quality variations. This could lead to more accurate predictions and better decision-making.
- **Advanced Machine Learning Algorithms**: Exploring more complex machine learning algorithms, such as neural networks or ensemble methods, could potentially enhance the accuracy of quality predictions by capturing intricate patterns in the data.

- **Real-time Visualization Dashboard**: Developing a real-time dashboard that visualizes the quality predictions and process performance could offer operators an intuitive tool to monitor and respond to variations effectively.
- **Collaboration with Domain Experts**: Collaborating with domain experts from the mining industry could provide deeper insights into specific factors influencing quality variations, leading to more tailored and accurate prediction models.
- **Data Augmentation Techniques**: Exploring data augmentation techniques, such as synthetic data generation, could help mitigate challenges posed by data scarcity and improve model robustness.